

## Modelação e estimação de parâmetros de acontecimentos raros, em ambiente R

### — as abordagens paramétrica e semi-paramétrica

(Poster)

Helena PENALVA – *Escola Superior de Ciências Empresariais/Instituto Politécnico de Setúbal*

helen.penalva@esce.ips.pt

Sandra Nunes – *Escola Superior de Ciências Empresariais /Instituto Politécnico de Setúbal e CMA*

Manuela Neves – *Instituto Superior de Agronomia/Universidade de Lisboa e CEAUL*

## Resumo

Historicamente, as aplicações da Teoria de Valores Extremos iniciaram-se em duas principais áreas: a área ambiental, com o estudo dos níveis do mar, velocidade do vento, caudal dos rios, entre outros; e a área da fiabilidade. Actualmente, a Teoria de Valores Extremos tem surgido como uma das mais importantes áreas da Estatística utilizada em várias ciências aplicadas, tais como, a Biologia, a Geologia e risco sísmico, a Climatologia e o Ambiente, ver Coles (2001) e Beirlant *et al.* (2004). A Teoria de Valores Extremos pretende estimar e prever acontecimentos que poderão ser mais extremos do que aqueles que alguma vez já foram observados. Assim, a inferência estatística somente pode ser deduzida a partir das observações que são consideradas extremas num determinado contexto. O estudo do comportamento dos extremos de amostras de variáveis aleatórias pode ser feito usando duas abordagens: a paramétrica e a semi-paramétrica. Iremos tratar neste trabalho apenas a análise estatística de extremos univariados. Em contexto paramétrico, o primeiro modelo de distribuições do máximo é o chamado *Método dos Máximos Anuais*, *Método dos Máximos de Blocos* ou ainda *Método de Gumbel*, que considera os  $k$  valores máximos de sub-amostras de tamanho  $r, n = k \times r$ . Uma outra abordagem é a que considera a modelação da distribuição de excessos acima de certo limiar, conhecida por *Método POT (Peaks Over Thresholds)*, na qual se restringe

a nossa atenção às observações que excedem um certo nível elevado,  $u$ . A abordagem semi-paramétrica teve a sua origem na escola holandesa com os trabalhos de de Haan (1970), e neste caso apenas se admite que o modelo  $F$  subjacente aos dados da amostra verifica certas propriedades na cauda. Nestas distribuições, o parâmetro de forma,  $\gamma$ , designado por índice de cauda ou índice de valores extremos descreve o comportamento da cauda direita,  $1 - F$ , do modelo subjacente aos dados. A sua estimação precisa é muito importante e de enorme influência na estimação de outros parâmetros, tais como quantis elevados ou período de retorno de quantis elevados. Neste trabalho pretende-se abordar a utilização do software R, (R Development Core Team, 2012) na estimação de parâmetros de acontecimentos raros, aplicando as abordagens paramétrica e semi-paramétrica. Para isso iremos escolher dois conjuntos de dados: um já tratado na literatura e outro conjunto de dados relativos a níveis médios diários na estação hidrométrica do rio Fraga.

**Agradecimentos:** Investigação parcialmente suportada por Fundos nacionais através da FCT—Fundação para a Ciência e a Tecnologia, projecto PEst-OE/MAT/UI0006/2011 e PEst-OE/MAT7UI0297/2011(CMA/FCT/UNL).

## Referências

- [1] Beirlant, J., Goegebeur, Y., Teugels, J. and Segers, J. (2004) *Statistics of Extremes: Theory and Applications*. Wiley, England.
- [2] Coles, S. (2001) *An Introduction to Statistical Modeling of extreme Values*. Springer-Verlag, London.
- [3] Haan, L.de (1970) *On regular variation and its Applications to the Weak Convergence of Sample Extremes*. Mathematical Centre Tract 32, Amsterdam.
- [4] R Development Core Team (2012). R: A Language and Environment for Statistical Computing, R. Foundation for Statistical Computing, Vienna, Austria, URL <http://www.R-project.org/>. ISBN 3-900051-07-0.